PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 13/00, C12N 1/21, 5/16
C12N 9/10, 15/12, 15/54
C12N 15/63, 15/79, C12P 21/00
C12Q 1/68

A1

(11) International Publication Number: WO 94/07917

(43) International Publication Date: 14 April 1994 (14.04.94)

(21) International Application Number: PCT/US93/09303

(22) International Filing Date: 30 September 1993 (30.09.93)

(30) Priority data:
07/955,041 1 October 1992 (01.10.92) US

(71) Applicant: LA JOLLA CANCER RESEARCH FOUNDATION [US/US]; 10901 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: FUKUDA, Minoru ; 2818 Passy Avenue, San Diego, CA 92122 (US). BIERHUIZEN, Marti, F., A. ; Schiedamesegeg 225, NL-3119 JB Schiedam (NL).

(74) Agents: KONSKI, Antoinette, F. et al.; Campbell & Flores, 4370 La Jolla Village Drive, Suite 700, San Diego, CA 92122 (US).

(81) Designated States: AU, CA, CZ, FI, HU, JP, KR, NO, NZ, RU, SK, UA.

Published
With international search report.

(54) Title: A NOVEL β1->6 *N*-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE, LEUKOSIALIN, AND A METHOD FOR CLONING PROTEINS HAVING ENZYMATIC ACTIVITY

(57) Abstract

The present invention provides a novel β1->6 *N*-acetylglucosaminyltransferase, which forms core 2 oligosaccharide structures in *O*-glycans, and a novel acceptor molecule, leukosialin, CD43, for core 2 β1->6 *N*-acetylglucosaminyltransferase activity. The amino acid sequences and nucleic acid sequences encoding these molecules, as well as active fragments thereof, also are disclosed. A method for isolating nucleic acid sequences encoding proteins having enzymatic activity is disclosed, using CHO cells that support replication of plasmid vectors having a polyoma virus origin of replication. A method to obtain a suitable cell line that expresses an acceptor molecule also is disclosed.

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 359..428
(D) OTHER INFORMATION: /note= "EXON 1 IS LOCATED IN GENOMIC DNA"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 193..806
(D) OTHER INFORMATION: /note= "THIS SEGMENT OF NUCLEIC ACID CONSTITUTES INTRON SEQUENCE OF THE cDNA"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 807..900
(D) OTHER INFORMATION: /note= "EXON 2 IS LOCATED IN BOTH GENOMIC AND cDNA. IN THE cDNA EXON 2 IMMEDIATELY FOLLOWS EXON 1'."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TTGGGGACCA CAAATGCAAA GGAAACCACC CTCCCCTCCC ACCTCCTCCT CTGCACCCTT     60

GAGTTCTCAG GCTCACATTC CCACCACCCA CCTCTGAGCC CAGCCCTCCC TAGCATCACC    120

ACTTCCATCC CATTCCTCAG CCAAGAGCCA GGAATCCTGA TTCCAGATCC CACGCTTCCC    180

TGCCTCCCTC AGGTGAGCCC CAGACCCCCA GGCACCCCGC TGGCCCCTGA AGGAGCAGGT    240

GATGGTGCTG TCTTCGCCCA GCAGCTGTGG GAGCAGGCGG GTGGGGCAGG ATGGAGGGGT    300

GGGTGGGGTG GGTGGAGCCA GGGCCCACTT CCTTTCCCCT TGGGGCCCTG TCCTTCCCAG    360

TCTTGCCCCA GCCTCGGGAG GTGGTGGAGT GACCTGGCCC CAGTGCTGCG TCCTTATCAG    420

CCGAGCCGGT AAGAGGGTGA GACTTGGTGG GGTAGGGGCC TCAGTGGGCC TGGGAATGTG    480

CCTGTGGCTT GAAAAGACTC TGACAGGTTA TGATGGGAAG AGATTGGGAG CCATTGGGCT    540

GCACAGGGTC AGGGAAGGCC AGGAGGGGCT GGTCACTGCT GGAATCTAAG CTGCTGAGGC    600

TGGAGGGAGC CTCAGGATGG GGCTGATGGG GGAGCTGCCA GCATCTGTTC CTCTGTCATT    660

TCTGATAACA GTAAAAGCCA GCATGGAAAA AACCGTTAAA CCGCAGGTTG GGCCTGGCCG    720

TTGGCAGGGA AGTGGGCAGA GGGGAGGCCC GGCCAGGTCC TCCGGCAACT CCCGCGTGTT    780

CTGCTTCTCC GGCTGCCCAC CTGCAGGTCC CAGCTCTTGC TCCTGCCTGT TTGCCTGGAA    840

ATG GCC ACG CTT CTC CTT CTC CTT GGG GTG CTG GTG GTA AGC CCA GAC       888
Met Ala Thr Leu Leu Leu Leu Leu Gly Val Leu Val Val Ser Pro Asp
  1               5                  10                  15

GCT CTG GGG AGC                                                      900
Ala Leu Gly Ser
            20
```

A NOVEL ß1→6 *N*-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE, LEUKOSIALIN, AND A METHOD FOR CLONING PROTEINS HAVING ENZYMATIC ACTIVITY

This work was supported by grants CA33000 and CA33895 awarded by the National Cancer Institute. The United States Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

This invention relates generally to the fields of biochemistry and molecular biology and more specifically to a novel human enzyme, UDP-GlcNAc:Galß1→3GalNAc (GlcNAc to GalNAc) ß1→6 *N*-acetylglucosaminyltransferase (core 2 ß1→6 *N*-acetylglucosaminyltransferase; C2GnT), and to a novel acceptor molecule, leukosialin, CD43, for core 2 ß1→6 *N*-acetylglucosaminyltransferase action. The invention additionally relates to DNA sequences encoding core 2 ß1→6 *N*-acetylglucosaminyltransferase and leukosialin, to vectors containing a C2GnT DNA sequence or a leukosialin DNA sequence, to recombinant host cells transformed with such vectors and to a method of transient expression cloning in CHO cells for identifying and isolating DNA sequences encoding specific proteins, using CHO cells expressing a suitable acceptor molecule.

BACKGROUND INFORMATION

Most *O*-glycosidic oligosaccharides in mammalian glycoproteins are linked via *N*-acetylgalactosamine to the hydroxyl groups of serine or threonine. These *O*-glycans can be classified into 4 different groups depending on the nature of the core portion of the oligosaccharides (see Fig. 1). Although less well studied than *N*-glycans, *O*-glycans likely have important biological functions. Indeed, the presence of *O*-linked oligosaccharides with the

(1985) reported T305 binding was abolished by neuraminidase treatment, suggesting T305 binds to hexasaccharides. T305 specifically reacts with the high molecular weight form of leukosialin (Saitoh et al., supra, (1991)).

Previous studies indicated poly-*N*-acetyllactosamine repeats extend almost exclusively from the branch formed by the core 2 ß1→6 *N*-acetylglucosaminyltransferase (Fukuda et al., J. Biol. Chem. 261:12796-12806 (1986)). Consistent with these results, Yousefi et al., supra, (1991) demonstrated that the core 2 enzyme in metastatic tumor cells regulates the level of poly-*N*-acetyllactosamine synthesis in *O*-linked oligosaccharides.

Poly-*N*-acetyllactosamines are subject to a variety of modifications, including the formation of the sialyl Le^x, NeuNAcα2→3Galß1→4(Fucα1→3)GlcNAc-, or the sialyl Le^a, NeuNAcα2→3Galß1→3 (Fucα1→4)GlcNAc-, determinants (Fukuda, Biochim. Biophys. Acta 780:119-150 (1985)). Such modifications are significant because these determinants, which are present on neutrophils and monocytes, serve as ligands for E- and P-selectin present on endothelial cells and platelets, respectively (see, for example, Larsen et al., Cell 63:467-474 (1990)).

In addition, tumor cells often express a significant amount of sialyl Le^x and/or sialyl Le^a on their cell surfaces. The interaction between E-selectin or P-selectin and these cell surface carbohydrates may play a role in tumor cell adhesion to endothelium during the metastatic process (Walz et al., supra, (1990)). Kojima et al., Biochem. Biophys. Res. Commun. 182:1288-1295 (1992) reported that selectin-dependent tumor cell adhesion to endothelial cells was abolished by blocking *O*-glycan synthesis. Complex sulfated *O*-glycans also may serve as ligands for the lymphocyte homing receptor, L-selectin

The invention further relates to a novel purified acceptor molecule, leukosialin, CD43, for core 2 ß1→6 *N*-acetylglucosaminyltransferase activity. The leukosialin cDNA encodes a novel variant leukosialin, which is created by alternative splicing of the genomic leukosialin DNA sequence.

Isolated nucleic acids encoding either core 2 ß1→6 *N*-acetylglucosaminyltransferase or leukosialin are disclosed, as are vectors containing the nucleic acids and recombinant host cells transformed with such vectors. The invention further provides methods of detecting such nucleic acids by contacting a sample with a nucleic acid probe having a nucleotide sequence capable of hybridizing with the isolated nucleic acids of the present invention. The core 2 ß1→6 *N*-acetylglucosaminyltransferase and leukosialin amino acid and nucleic acid sequences disclosed herein can be purified from human cells or produced using well known methods of recombinant DNA technology.

The invention also discloses a method of isolating nucleic acid sequences encoding proteins that have an enzymatic activity. Such a nucleic acid sequence is obtained by transfecting the nucleic acid, which is contained within a vector having a polyoma virus replication origin, into a Chinese hamster ovary (CHO) cell line simultaneously expressing polyoma virus large T antigen and the acceptor molecule for the protein having an enzymatic activity.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the structures and biosynthesis of *O*-glycans. Structures of *O*-glycan cores can be classified into 4 groups (core 1 to core 4), each of which is synthesized starting with GalNAcα1→Ser/Thr. The core 1 structure is synthesized by the addition of a ß1→3 Gal

1). Plasmid DNA was extracted using the Hirt procedure and samples were digested with XhoI and DpnI. In parallel, pGT/hCG plasmid purified from *E. coli* MC1061/P3 was digested with XhoI and DpnI (lane 7 in panel A and lane 4 in panel B) or XhoI alone (lane 8 in panel A and lane 5 in panel B). The arrow indicates the migration of plasmid DNA resistant to DpnI digestion. The arrowheads indicate plasmid DNA digested by DpnI.

Figure 4 shows the expression of T305 antigen expressed by pcDNAI-C2GnT. Subconfluent CHO-Py-leu cells were transfected with pcDNAI-C2GnT (panels A and B) or mock-transfected with pcDNAI (panels C and D). Sixty four hours after transfection, the cells were fixed, then incubated with mouse T305 monoclonal antibody followed by fluorescein isocyanate-conjugated sheep anti-mouse IgG (panels A, B and C). Two different areas are shown in panels A and B. Panel D shows a phase micrograph of the same field shown in panel C. Bar = 20μm.

Figure 5 depicts the cDNA sequence (SEQ. ID. NO. 4) and translated amino acid sequences (SEQ. ID. NO. 5) of core 2 ß1→6 *N*-acetylglucosaminyltransferase. The open reading frame and full-length nucleotide sequence of C2GnT are shown. The signal/membrane-anchoring domain is doubly underlined. The polyadenylation signal is boxed. Potential *N*-glycosylation sites are marked with asterisks. The sequences are numbered relative to the translation start site.

Figure 6 shows the expression of core 2 ß1→6 *N*-acetylglucosaminyltransferase mRNA in various cell types. Poly(A)$^+$ RNA (11 μg) from CHO-Py-leu cells (lane 1), HL-60 promyelocytes (lane 2), K562 erythrocytic cells (lane 3), and SP and L4 colonic carcinoma cells (lanes 4 and 5) was resolved by electrophoresis. RNA was transferred to a nylon membrane and hybridized with a radiolabeled fragment

encoding a glycosyltransferase requires an appropriate recipient cell line. Ideal recipient cells should not express the glycosyltransferase of interest. As a result, the recipient cells would normally lack the oligosaccharide structure formed by such a glycosyltransferase.

Expression of the cloned glycosyltransferase cDNA in the recipient cell line should result in formation of the specific oligosaccharide structure. The resultant oligosaccharide can be identified using a specific antibody or lectin that recognizes the structure. The recipient cell line also must support replication of an appropriate plasmid vector.

COS-1 cells initially appear to satisfy the requirements for using the transient expression method. COS-1 cells express SV40 large T antigen and support the replication of plasmid vectors harboring a SV40 replication origin (Gluzman et al., Cell 23:175-182 (1981)). Although COS-1 cells, themselves, express a variety of glycosyltransferases, COS-1 cells have been used to clone cDNA sequences encoding human blood group Lewis α1→3/4 fucosyltransferase and murine α1→3 galactosyltransferase (Kukowska-Latallo et al., Genes and Devel. 4:1288-1303 (1990); Larsen et al., Proc. Natl. Acad. Sci. *USA* 86:8227-8231 (1989)). Also, Goelz et al., Cell 63:175-182 (1990), utilized an antibody that inhibits E-selectin mediated adhesion to isolate a cDNA sequence encoding α1→3 fucosyltransferase.

An attempt was made to use COS-1 cells to isolate cDNA clones encoding core 2 ß1→6 *N*-acetylglucosaminyltransferase. COS-1 cells were transfected using cDNA obtained from activated human T cells, which express the core 2 ß1→6 *N*-acetylglucosaminyltransferase. Transfected cells suspected of expressing core 2 ß1→6 *N*-acetylglucosaminyltransferase

insert was sequenced (see Figure 5; SEQ. ID. NO. 4). The 2105 base pair cDNA sequence encodes a putative 428 amino acid protein. The genomic DNA sequence encoding can be isolated using methods well known to those skilled in the art, such as nucleic acid hybridization using the core 2 ß1→6 *N*-acetylglucosaminyltransferase cDNA disclosed herein to screen, for example, a genomic library prepared from HL-60 promyelocytes.

An enzyme similar to the disclosed human core 2 ß1→6 *N*-acetylglucosaminyltransferase has been purified from bovine tracheal epithelium (Ropp et al., J. Biol. Chem. 266:23863-23871 (1991), which is incorporated herein by reference. The apparent molecular weight of the bovine enzyme is ~69kDa. In comparison, the predicted molecular weight of the polypeptide portion of core 2 ß1→6 *N*-acetylglucosaminyltransferase is ~50kDa. The deduced amino acid sequence of core 2 ß1→6 *N*-acetylglucosaminyltransferase reveals two to three potential *N*-glycosylation sites, suggesting *N*-glycosylation and *O*-glycosylation, or other post-translational modification, could account for the larger apparent size of the bovine enzyme.

Expression of the cloned C2GnT sequence, or a fragment thereof, directed formation of the specific *O*-glycan core 2 oligosaccharide structure. Although several cDNA sequences encoding glycosyltransferases have been isolated (Paulson and Colley, J. Biol. Chem. 264:17615-17618 (1989); Schachter, Curr. Opin. Struct. Biol. 1:755-765 (1991), which are incorporated herein by reference), C2GnT is the first reported cDNA sequence encoding an enzyme involved exclusively in *O*-glycan synthesis.

In *O*-glycans, ß1→6 *N*-acetylglucosaminyl linkages may occur in both core 2, Galß1→3(GlcNAcß1→6)GalNAc, and core 4, GlcNAcß1→3(GlcNAcß1→6)GalNAc, structures

The acceptor molecule specificity of core 2 ß1→6 *N*-acetylglucosaminyltransferase is different from the specificity of the enzymes present in tracheal epithelium and Novikoff hepatoma cells. Thus, a family of ß1→6 *N*-acetylglucosaminyltransferases can exist, the members of which differ in acceptor specificity but are capable of forming the same linkage. Members of this family are isolated from cells expressing ß1→6 *N*-acetylglucosaminyltransferase activity, using, for example, nucleic acid hybridization assays and studies of acceptor molecule specificity. Such a family was reported for the α1→3 fucosyltransferases (Weston et al., J. Biol. Chem. 267:4152-4160 (1992), which is incorporated herein by reference).

The formation of the core 2 structure is critical to cell structure and function. For example, the core 2 structure is essential for elongation of poly-*N*-acetyllactosamine and for formation of sialyl Le^x or sialyl Le^a structures. Furthermore, the biosynthesis of cartilage keratan sulfate may be initiated by the core 2 ß1→6 *N*-acetylglucosaminyltransferase, since the keratan sulfate chain is extended from a branch present in core 2 structure in the same way as poly-*N*-acetyllactosamine (Dickenson et al., Biochem. J. 269:55-59 (1990), which is incorporated herein by reference). Keratan sulfate is absent in wild-type CHO cells, which do not express the core 2 ß1→6 *N*-acetylglucosaminyltransferase (Esko et al., J. Biol. Chem. 261:15725-15733 (1986), which is incorporated herein by reference). These structures are believed to be important for cellular recognition and matrix formation. The availability of the cDNA clone encoding the core 2 ß1→6 *N*-acetylglucosaminyltransferase will aid in understanding how the various carbohydrate structures are formed during differentiation and malignancy. Manipulation of the expression of the various carbohydrate structures by gene

acetylglucosaminyltransferase antibodies can be used to substantially purify naturally-occurring core 2 ß1→6 *N*-acetylglucosaminyltransferase from human HL-60 promyelocytes.

Alternatively, a purified protein of the present invention can be obtained by well known recombinant methods, utilizing the nucleic acids disclosed herein, as described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d ed. (Cold Spring Harbor Laboratory 1989), which is incorporated herein by reference, and by the methods described in the Examples below. Furthermore, purified proteins can be synthesized by methods well known in the art.

As used herein, the phrase "substantially the sequence" includes the described nucleotide or amino acid sequence and sequences having one or more additions, deletions or substitutions that do not substantially affect the ability of the sequence to encode a protein have a desired functional activity. In addition, the phrase encompasses any additional sequence that hybridizes to the disclosed sequence under stringent hybridization sequences. Methods of hybridization are well known to those skilled in the art. For example, sequence modifications that do not substantially alter such activity are intended. Thus, a protein having substantially the amino acid sequence of Figure 5 (SEQ. ID. NO. 5) refers to core 2 ß1→6 *N*-acetylglucosaminyltransferase encoded by the cDNA described in Example IV, as well as proteins having amino acid sequences that are modified but, nevertheless, retain the functions of core 2 ß1→6 *N*-acetylglucosaminyltransferase. One skilled in the art can readily determine such retention of function following the guidance set forth, for example, in Examples V and VI.

As used herein, the term "critical branches" refers to oligosaccharide structures formed by specific glycosyltransferase activity. Critical branches may be involved in various cellular functions, such as cell-cell recognition. The oligosaccharide structure of a critical branch can be determined using methods well known in the art, such as the method for determining the core 2 oligosaccharide structure, as described in Examples V and VI.

Relatedly, the invention also provides nucleic acids encoding the human core 2 β1→6 *N*-acetylglucosaminyltransferase protein and leukosialin protein described above. The nucleic acids can be in the form of DNA, RNA or cDNA, such as the novel C2GnT cDNA of 2105 base pairs identified in Figure 5 (SEQ. ID. NO. 4) or the novel leukosialin cDNA identified in Figure 2 (SEQ. ID. NO. 2), for example. Such nucleic acids can also be chemically synthesized by methods known in the art, including, for example, the use of an automated nucleic acid synthesizer.

The nucleic acid can have substantially the nucleotide sequence of C2GnT, identified in Figure 5 (SEQ. ID. NO. 4), or leukosialin identified in Figure 2 (SEQ. ID. NO. 2). Portions of such nucleic acids that encode active fragments of the core 2 β1→6 *N*-acetylglucosaminyltransferase protein or leukosialin protein of the present invention also are contemplated.

Nucleic acid probes capable of hybridizing to the nucleic acids of the present invention under reasonably stringent conditions can be prepared from the cloned sequences or by synthesizing oligonucleotides by methods known in the art. The probes can be labeled with markers according to methods known in the art and used to detect the nucleic acids of the present invention. Methods for

sequences, the C2Gnt DNA sequence or the leukosialin DNA sequence, and an antibiotic resistance gene for selection. The construct can then be transfected into a suitable cell line, such as PA12, which carries a packaging deficient provirus and expresses the necessary components for virus production, including synthesis of amphotrophic glycoproteins. The supernatant from these cells contain infectious virus, which can be used to infect the cells of interest.

Isolated recombinant polypeptides or proteins can be obtained by growing the described host cells under conditions that favor transcription and translation of the transfected nucleic acid. Recombinant proteins produced by the transfected host cells are isolated using methods set forth herein and by methods well known to those skilled in the art.

Also provided are antibodies having specific reactivity with the core 2 ß1→6 *N*-acetylglucosaminyltransferase protein or leukosialin protein of the present invention. Active fragments of antibodies, for example, Fab and Fab'$_2$ fragments, having specific reactivity with such proteins are intended to fall within the definition of an "antibody." Antibodies exhibiting a titer of at least about 1.5 x 10^5, as determined by ELISA, are useful in the present invention.

The antibodies of the invention can be produced by any method known in the art. For example, polyclonal and monoclonal antibodies can be produced by methods described in Harlow and Lane, Antibodies: A Laboratory Manual (Cold Spring Harbor 1988), which is incorporated herein by reference. The proteins, particularly core 2 ß1→6 *N*-acetylglucosaminyltransferase or leukosialin of the present invention can be used as immunogens to generate such antibodies. Altered antibodies, such as chimeric,

EXAMPLE I

EXPRESSION CLONING IN COS-1 CELLS OF THE cDNA FOR THE PROTEIN CARRYING THE HEXASACCHARIDES

COS-1 cells were transfected with a cDNA library, pcDSRα-2F1, constructed from poly(A)$^+$ RNA of activated T lymphocytes, which express the core 2 ß1→6 *N*-acetylglucosaminyltransferase (Yokota et al., Proc. Natl. Acad. Sci. *USA* 83:5894-5898 (1986); Piller et al., supra, (1988), which are incorporated herein by reference). COS-1 cells support replication of the pcDSRα constructs, which contain the SV40 replication origin. Transfected cells were selected by panning using monoclonal antibody T305, which recognizes sialylated branched hexasaccharides (Piller et al., supra, (1991); Saitoh et al., supra, (1991)). Methods referred to in this example are described in greater detail in the examples that follow.

Following several rounds of transfection, one plasmid, pcDSRα-leu, directing high expression of the T305 antigen was identified. The cloned cDNA insert was isolated and sequenced, then compared with other reported sequences. The newly isolated cDNA sequence was nearly identical to the sequence reported for leukosialin, except the 5′-flanking sequences were different (Pallant et al., Proc. Natl. Acad. Sci. *USA* 86:1328-1332 (1989), which is incorporated herein by reference).

Comparison of the cloned cDNA sequence with the genomic leukosialin DNA sequence revealed the start site of the cDNA sequence is located 259 bp upstream of the transcription start site of the previously reported sequence (Figure 2; compare Exon 1′ and Exon 1) (Shelley et al., Biochem. J. 270:569-576 (1990); Kudo and Fukuda, J. Biol. Chem. 266:8483-8489 (1991), which are incorporated herein by reference). A consensus splice site was

cDNA clone expressing core 2 ß1→6 *N*-acetylglucosaminyltransferase activity, a CHO cell line expressing both leukosialin and polyoma large T antigen was established (see, for example, Heffernan and Dennis Nucl. Acids Res. 19:85-92 (1991), which is incorporated herein by reference).

Vectors: A plasmid vector, pPSVE1-PyE, which contains the polyoma virus early genes under the control of the SV40 early promoter, was constructed using a modification of the method of Muller et al., Mol. Cell. Biol. 4:2406-2412 (1984), which is incorporated herein by reference. Plasmid pPSVE1 was prepared using pPSG4 (American Type Culture Collection 37337) and SV40 viral DNA (Bethesda Research laboratories) essentially as described by Featherstone et al., Nucl. Acids Res. 12:7235-7249 (1984), which is incorporated herein by reference. Following EcoRI and HincII digestion of plasmid pPyLT-1 (American Type Culture Collection 41043), a DNA sequence containing the carboxy terminal coding region of polyoma virus large T antigen was isolated. The HincII site was converted to an EcoRI site by blunt-end ligation of phosphorylated EcoRI linkers (Stratagene). Plasmid pPSVE1-PyE was generated by inserting the carboxy-terminal coding sequence for large T antigen into the unique EcoRI site of plasmid pPSVE1.

Plasmid pZIPNEO-leu was constructed by introducing the EcoRI fragment of PEER-3 cDNA, which contains the complete coding sequence for human leukosialin, into the unique EcoRI site of plasmid pZIPNEO (Cepko et al., Cell 37:1053-1063 (1984), which is incorporated herein by reference). Plasmid structures were confirmed by restriction mapping and by sequencing the construction sites. pZIPNEO was kindly provided by Dr. Channing Der.

Transfection: CHODG44 cells were grown in 100 mm tissue

site, "GATC". The methylated DpnI recognition site is susceptible to cleavage by DpnI. In contrast, the DpnI recognition site of plasmids replicated in mammalian cells is not methylated and, therefore, is resistant to DpnI digestion.

Methylated plasmid pGT/hCG was transfected by lipofection into each of the six selected clonal cell lines expressing leukosialin. After 64 hr, low molecular weight plasmid DNA was isolated from the cells using the method of Hirt, J. Mol. Biol. 26:365-369 (1967), which is incorporated herein by reference. Isolated plasmid DNA was digested with XhoI and DpnI (Stratagene), subjected to electrophoresis in a 1% agarose gel, and transferred to nylon membranes (Micron Separations Inc., MA).

A 0.4 kb SmaI fragment of the ß1→4 galactosyltransferase DNA sequence of pGT/hCG was radiolabeled with [^{32}P]dCTP using the random primer method (Feinberg and Vogelstein, Anal. Biochem. 132:6-13 (1983), which is incorporated herein by reference). Hybridization was performed using methods well-known to those skilled in the art (see, for example, Sambrook et al., supra, (1989)). Following hybridization, the membranes were washed several times, including a final high stringency wash in 0.1 x SSPE, 0.1% SDS for 1 hr at 65°C, then exposed to Kodak X-AR film at -70°C.

Four of the six clones tested supported replication of the pcDNAI-based plasmid, pGT/hCG (Fig. 3.A., lanes 1, 3, 4 and 5). MOP-8 cells, a 3T3 cell line transformed by polyoma virus early genes (Muller et al., supra, (1984)), expresses endogenous core 2 ß1→6 *N*-acetylglucosaminyltransferase activity and was used as a control for the replication assay (Fig. 3.B., lane 1). One clonal cell line that supported pGT/hCG replication, CHO-Py-leu (Fig. 3.A., lane 5; Fig. 3.B., lanes 2 and 3) and

EDTA/5% fetal calf serum, pH7.4, containing a 1:200 dilution of ascites fluid containing T305 monoclonal antibody. The cells were incubated on ice for 1 hr, then washed in the same buffer and panned on dishes coated with goat anti-mouse IgG (Sigma) (Wysocki and Sato Proc. Natl. Acad. Sci. *USA* 75:2844-2848 (1978); Seed & Aruffo Proc. Natl. Acad. Sci. *USA* 84:3365-3369 (1987), which are incorporated herein by reference). T305 monoclonal antibody was kindly provided by Dr. R.I. Fox, Scripps Research Foundation, La Jolla, CA.

Plasmid DNA was recovered from adherent cells by the method of Hirt, supra, (1967), treated with DpnI to eliminate plasmids that had not replicated in transfected cells, and transformed into *E. coli* strain MC1061/P3. Plasmid DNA was then recovered and subjected to a second round of screening. *E. coli* transformants containing plasmids recovered from this second enrichment were plated to yield 8 pools of approximately 500 colonies each. Replica plates were prepared using methods well-known to those skilled in the art (see, for example, Sambrook et al., supra, (1989)).

The pooled plasmid DNA was prepared from replica plates and transfected into CHO-Py-leu cells. The transfectants were screened by panning. One plasmid pool was selected and subjected to three subsequent rounds of selection. One plasmid, pcDNAI-C2GnT, which directed the expression of the T305 antigen, was isolated. CHO-Py-leu cells transfected with pcDNAI-C2GnT express the antigen recognized by T305, whereas CHO-Py-leu cells transfected with pcDNAI are negative for T305 antigen (Fig. 4). These results show pcDNAI-C2GnT directs the expression of a new determinant on leukosialin that is recognized by T305 monoclonal antibody. This determinant is the branched hexasaccharide sequence, NeuNAcα2→3Galß1→3(NeuNAcα2→3Galß1→4 GlcNAcß1→6)GalNAc.

The putative protein contains three potential *N*-glycosylation sites (Fig. 5, asterisks). However, one of these sites contains a proline residue adjacent to asparagine and is not likely utilized in vivo.

No matches were obtained when the C2GnT cDNA sequence and deduced amino acid sequence were compared with sequences listed in the PC/Gene 6.6 data bank. In particular, no homology was revealed between the deduced amino acid sequence of C2GnT and other glycosyltransferases, including *N*-acetylglucosaminyltransferase I (Sarkar et al., Proc. Natl. Acad. Sci. *USA* 88:234-238 (1991), which is incorporated herein by reference).

mRNA expression: Poly(A)$^+$ RNA was prepared using a kit (Stratagene) and resolved by electrophoresis on a 1.2% agarose/2.2 M formaldehyde gel, and transferred to nylon membranes (Micro Separations Inc., MA) using methods well-known to those skilled in the art (see, for example, Sambrook et al., supra, (1989)). Membranes were probed using the EcoRI insert of pPROTA-C2GnT (see below) radiolabeled with [^{32}P]dCTP by the random priming method (Feinberg and Vogelstein, supra, (1983). Hybridization was performed in buffers containing 50% formamide for 24 hr at 42°C (Sambrook et al., supra, (1989)). Following hybridization, filters were washed several times in 1xSSPE/0.1% SDS at room temperature and once in 0.1xSSPE/0.1% SDS at 42°C, then exposed to Kodak X-AR film at -70°C.

Fig. 6 compares the level of core 2 ß1→6 *N*-acetylglucosaminyltransferase mRNA isolated from HL-60 promyelocytes, K562 erythroleukemia cells, and poorly metastatic SP and highly metastatic L4 colonic carcinoma cells. The major RNA species migrates at a size essentially identical to the ~2.1 kb C2GnT cDNA sequence.

Reactions were incubated for 1 hr at 37°C, then processed by C18 Sep-Pak chromatography (Waters) (Palcic et al., J. Biol. Chem. 265:6759-6769 (1990), which is incorporated herein by reference). Core 2 and core 4 ß1→6 *N*-acetylglucosaminyltransferase were assayed using the acceptors p-nitrophenyl Galß1→3GalNAc and p-nitrophenyl GlcNAcß1→3GalNAc, respectively (Toronto Research Chemicals).

UDP-GlcNAc:α-Man ß1→6 *N*-acetylglucosaminyltransferase(V) was assayed using the acceptor GlcNAcß1→2Manα1→6Glc-ß-O-$(CH_2)_7CH_3$. The blood group I enzyme, UDP-GlcNAc:GlcNAcß1→3Galß1→4GlcNAc (GlcNAc to Gal) ß1→6 *N*-acetylglucosaminyltransferase, was assayed using GlcNAcß1→3Galß1→4GlcNAcß1→6Manα1→6Manß1→O-$(CH_2)_8COOCH_3$ or Galß1→4GlcNAcß1→3Galß1→4GlcNAcß1→3Galß1→4GlcNAcß1→O-$(CH_2)_7CH_3$ as acceptors (Gu et al., J. Biol. Chem. 267:2994-2999 (1992), which is incorporated herein by reference). Synthetic acceptors were kindly provided by Dr. Olé Hindsgaul, University of Alberta, Canada.

Results of these assays are shown in Table I. Assuming transfection efficiency of the cells is approximately 20-30%, the level of enzymatic activity directed by cells transfected with pcDNAI-C2GnT is roughly equivalent to the level observed in HL-60 cells.

(EcoRI recognition sites underlined). The EcoRI sites allowed direct, in-frame insertion of the fragment into the unique EcoRI site of plasmid pPROTA (Sanchez-Lopez et al., J. Biol. Chem. 263:11892-11899 (1988), which is incorporated herein by reference).

The nucleotide sequence of the insert as well as the proper orientation were confirmed by DNA sequencing using the primers described above for cDNA sequencing. Plasmid pPROTA-C2GnT allows secretion of the fusion protein from transfected cells and binding of the secreted fusion protein by insolubilized immunoglobulins.

Either pPROTA or pPROTA-C2GnT was transfected into COS-1 cells. Following a 64 hr period to allow transient expression, cell supernatants were collected (Kukowska-Latallo et al., supra, (1990)). Cell supernatants were cleared by centrifugation, adjusted to 0.05% Tween 20 and either assayed directly for core 2 ß1→6 *N*-acetylglucosaminyltransferase activity or used in IgG-Sepharose (Pharmacia) binding studies. For the latter assay, supernatants (10 ml) were incubated batchwise with approximately 300 µl of IgG-Sepharose for 4 hr at 4°C. The matrices were then extensively washed and used directly for glycosyltransferase assays.

No core 2 ß1→6 *N*-acetylglucosaminyltransferase activity was detected in the medium of COS-1 cells transfected with the control plasmid, pPROTA. Similarly, no enzymatic activity was associated with IgG-Sepharose beads. In contrast, a significant level of core 2 ß1→6 *N*-acetylglucosaminyltransferase activity was detected in the medium of COS-1 cells transfected with pPROTA-C2GnT. The activity also associated with the IgG-Sepharose beads (Table II). No activity was detected in the supernatant following incubation of the supernatant with IgG-Sepharose.

EXAMPLE VI

DETERMINATION OF C2GnT SPECIFICITY

Four types of ß 1 → 6 *N*-acetylglucosaminyltransferase linkages have been reported, including core 2 and core 4 in *O*-glycans, I-antigen and a branch attached to mannose that forms tetraantennary *N*-glycans (see Table II). In order to determine whether these different structures are also synthesized by the cloned C2GnT cDNA sequence, enzymatic activity was determined using five different acceptors.

As shown in Table II, the fusion protein was only active with the acceptor for core 2 formation. The same was true when the formation of ß1→6 *N*-acetylglucosaminyl linkage to internal galactose residues was examined (Table II, see structure at bottom). This result precludes the likelihood that the enzyme encoded by the C2GnT cDNA sequence may add *N*-acetylglucosamine to a non-reducing terminal galactose. The HL-60 core 2 ß1→6 *N*-acetylglucosaminyltransferase is exclusively responsible for the formation of the GlcNAcß1→6 branch on Galß1→3 GalNAc.

Although the invention has been described with reference to the disclosed embodiments, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

```
AGT AGT GAT ATT AAT TGC ACC AAA GTT TTA CAG GGT GAT GTA AAT GAA      426
Ser Ser Asp Ile Asn Cys Thr Lys Val Leu Gln Gly Asp Val Asn Glu
     55                  60                  65

ATC CAA AAG GTA AAG CTT GAG ATC CTA ACA GTG AAA TTT AAA AAG CGC      474
Ile Gln Lys Val Lys Leu Glu Ile Leu Thr Val Lys Phe Lys Lys Arg
 70                  75                  80                  85

CCT CGG TGG ACA CCT GAC GAC TAT ATA AAC ATG ACC AGT GAC TGT TCT      522
Pro Arg Trp Thr Pro Asp Asp Tyr Ile Asn Met Thr Ser Asp Cys Ser
                 90                  95                 100

TCT TTC ATC AAG AGA CGC AAA TAT ATT GTA GAA CCC CTT AGT AAA GAA      570
Ser Phe Ile Lys Arg Arg Lys Tyr Ile Val Glu Pro Leu Ser Lys Glu
            105                 110                 115

GAG GCG GAG TTT CCA ATA GCA TAT TCT ATA GTG GTT CAT CAC AAG ATT      618
Glu Ala Glu Phe Pro Ile Ala Tyr Ser Ile Val Val His His Lys Ile
        120                 125                 130

GAA ATG CTT GAC AGG CTG CTG AGG GCC ATC TAT ATG CCT CAG AAT TTC      666
Glu Met Leu Asp Arg Leu Leu Arg Ala Ile Tyr Met Pro Gln Asn Phe
    135                 140                 145

TAT TGC GTT CAT GTG GAC ACA AAA TCC GAG GAT TCC TAT TTA GCT GCA      714
Tyr Cys Val His Val Asp Thr Lys Ser Glu Asp Ser Tyr Leu Ala Ala
150                 155                 160                 165

GTG ATG GGC ATC GCT TCC TGT TTT AGT AAT GTC TTT GTG GCC AGC CGA      762
Val Met Gly Ile Ala Ser Cys Phe Ser Asn Val Phe Val Ala Ser Arg
                170                 175                 180

TTG GAG AGT GTG GTT TAT GCA TCG TGG AGC CGG GTT CAG GCT GAC CTC      810
Leu Glu Ser Val Val Tyr Ala Ser Trp Ser Arg Val Gln Ala Asp Leu
            185                 190                 195

AAC TGC ATG AAG GAT CTC TAT GCA ATG AGT GCA AAC TGG AAG TAC TTG      858
Asn Cys Met Lys Asp Leu Tyr Ala Met Ser Ala Asn Trp Lys Tyr Leu
        200                 205                 210

ATA AAT CTT TGT GGT ATG GAT TTT CCC ATT AAA ACC AAC CTA GAA ATT      906
Ile Asn Leu Cys Gly Met Asp Phe Pro Ile Lys Thr Asn Leu Glu Ile
    215                 220                 225

GTC AGG AAG CTC AAG TTG TTA ATG GGA GAA AAC AAC CTG GAA ACG GAG      954
Val Arg Lys Leu Lys Leu Leu Met Gly Glu Asn Asn Leu Glu Thr Glu
230                 235                 240                 245

AGG ATG CCA TCC CAT AAA GAA GAA AGG TGG AAG AAG CGG TAT GAG GTC     1002
Arg Met Pro Ser His Lys Glu Glu Arg Trp Lys Lys Arg Tyr Glu Val
                250                 255                 260

GTT AAT GGA AAG CTG ACA AAC ACA GGG ACT GTC AAA ATG CTT CCT CCA     1050
Val Asn Gly Lys Leu Thr Asn Thr Gly Thr Val Lys Met Leu Pro Pro
            265                 270                 275

CTC GAA ACA CCT CTC TTT TCT GGC AGT GCC TAC TTC GTG GTC AGT AGG     1098
Leu Glu Thr Pro Leu Phe Ser Gly Ser Ala Tyr Phe Val Val Ser Arg
        280                 285                 290

GAG TAT GTG GGG TAT GTA CTA CAG AAT GAA AAA ATC CAA AAG TTG ATG     1146
Glu Tyr Val Gly Tyr Val Leu Gln Asn Glu Lys Ile Gln Lys Leu Met
    295                 300                 305

GAG TGG GCA CAA GAC ACA TAC AGC CCT GAT GAG TAT CTC TGG GCC ACC     1194
Glu Trp Ala Gln Asp Thr Tyr Ser Pro Asp Glu Tyr Leu Trp Ala Thr
310                 315                 320                 325
```

```
Arg Ile His Gln Lys Pro Glu Phe Val Ser Val Arg His Leu Glu Leu
         35                  40                  45

Ala Gly Glu Asn Pro Ser Ser Asp Ile Asn Cys Thr Lys Val Leu Gln
     50                  55                  60

Gly Asp Val Asn Glu Ile Gln Lys Val Lys Leu Glu Ile Leu Thr Val
 65                  70                  75                  80

Lys Phe Lys Lys Arg Pro Arg Trp Thr Pro Asp Asp Tyr Ile Asn Met
                 85                  90                  95

Thr Ser Asp Cys Ser Ser Phe Ile Lys Arg Arg Lys Tyr Ile Val Glu
            100                 105                 110

Pro Leu Ser Lys Glu Glu Ala Glu Phe Pro Ile Ala Tyr Ser Ile Val
        115                 120                 125

Val His His Lys Ile Glu Met Leu Asp Arg Leu Leu Arg Ala Ile Tyr
    130                 135                 140

Met Pro Gln Asn Phe Tyr Cys Val His Val Asp Thr Lys Ser Glu Asp
145                 150                 155                 160

Ser Tyr Leu Ala Ala Val Met Gly Ile Ala Ser Cys Phe Ser Asn Val
                165                 170                 175

Phe Val Ala Ser Arg Leu Glu Ser Val Val Tyr Ala Ser Trp Ser Arg
            180                 185                 190

Val Gln Ala Asp Leu Asn Cys Met Lys Asp Leu Tyr Ala Met Ser Ala
        195                 200                 205

Asn Trp Lys Tyr Leu Ile Asn Leu Cys Gly Met Asp Phe Pro Ile Lys
    210                 215                 220

Thr Asn Leu Glu Ile Val Arg Lys Leu Lys Leu Leu Met Gly Glu Asn
225                 230                 235                 240

Asn Leu Glu Thr Glu Arg Met Pro Ser His Lys Glu Glu Arg Trp Lys
                245                 250                 255

Lys Arg Tyr Glu Val Val Asn Gly Lys Leu Thr Asn Thr Gly Thr Val
            260                 265                 270

Lys Met Leu Pro Pro Leu Glu Thr Pro Leu Phe Ser Gly Ser Ala Tyr
        275                 280                 285

Phe Val Val Ser Arg Glu Tyr Val Gly Tyr Val Leu Gln Asn Glu Lys
    290                 295                 300

Ile Gln Lys Leu Met Glu Trp Ala Gln Asp Thr Tyr Ser Pro Asp Glu
305                 310                 315                 320

Tyr Leu Trp Ala Thr Ile Gln Arg Ile Pro Glu Val Pro Gly Ser Leu
                325                 330                 335

Pro Ala Ser His Lys Tyr Asp Leu Ser Asp Met Gln Ala Val Ala Arg
            340                 345                 350

Phe Val Lys Trp Gln Tyr Phe Glu Gly Asp Val Ser Lys Gly Ala Pro
        355                 360                 365

Tyr Pro Pro Cys Asp Gly Val His Val Arg Ser Val Cys Ile Phe Gly
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Asn Ser Pro Glu
1 5

10. The acceptor molecule of claim 9, wherein said acceptor molecule is leukosialin, CD43.

11. An isolated nucleic acid encoding the acceptor molecule of claim 9.

12. A vector containing the nucleic acid of claim 11.

13. The vector of claim 12, wherein said vector is a plasmid.

14. The vector of claim 12, wherein said vector is pcDSRα-leu.

15. A host cell containing the vector of claim 12.

16. A method of obtaining from a cell line, which does not normally contain a protein having catalytic activity or an acceptor molecule for said protein, a nucleic acid encoding said protein having catalytic activity comprising:

a. transfecting said cell line with a DNA sequence encoding the acceptor molecule, wherein the acceptor molecule is stably expressed in the cell line;

b. transfecting said cell line with a cDNA library containing said nucleic acid in a vector, wherein proteins encoded by the transfected cDNA are transiently expressed;

c. screening the transfected cells for expression of said protein having catalytic activity; and

d. isolating the nucleic acid encoding the protein having catalytic activity.

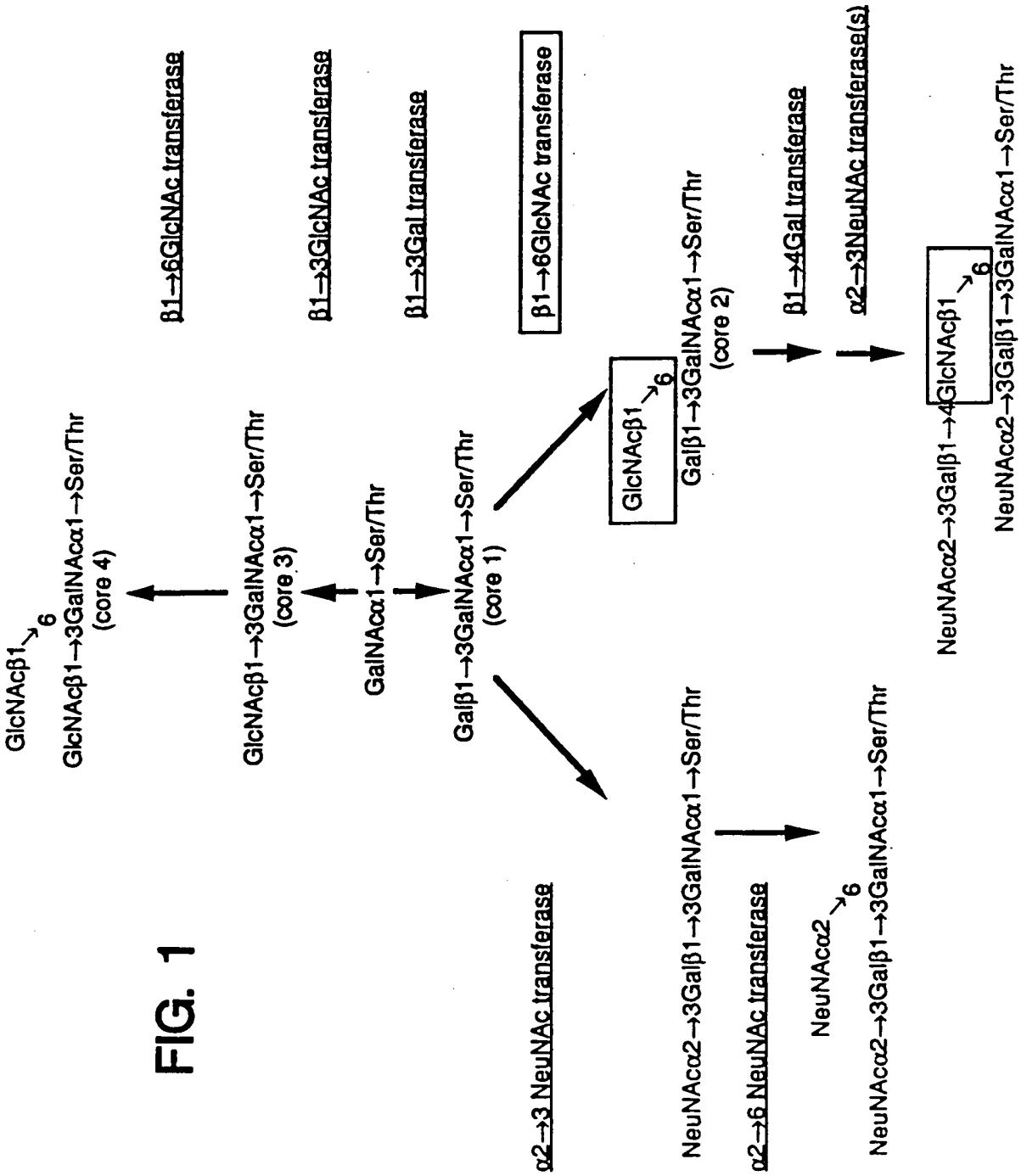
GlcNAcβ1→6 GlcNAcβ1→3GalNAcα1→Ser/Thr (core 4)
GlcNAcβ1→3GalNAcα1→Ser/Thr (core 3)
GalNAcα1→Ser/Thr
Galβ1→3GalNAcα1→Ser/Thr (core 1)
β1→6GlcNAc transferase
β1→3GlcNAc transferase
β1→3Gal transferase
β1→6GlcNAc transferase
α2→3 NeuNAc transferase
α2→6 NeuNAc transferase
GlcNAcβ1→6 Galβ1→3GalNAcα1→Ser/Thr (core 2)
NeuNAcα2→3Galβ1→3GalNAcα1→Ser/Thr
NeuNAcα2→6 NeuNAcα2→3Galβ1→3GalNAcα1→Ser/Thr
β1→4Gal transferase
α2→3NeuNAc transferase(s)
NeuNAcα2→3Galβ1→4GlcNAcβ1→6 NeuNAcα2→3Galβ1→3GalNAcα1→Ser/Thr

FIG. 1

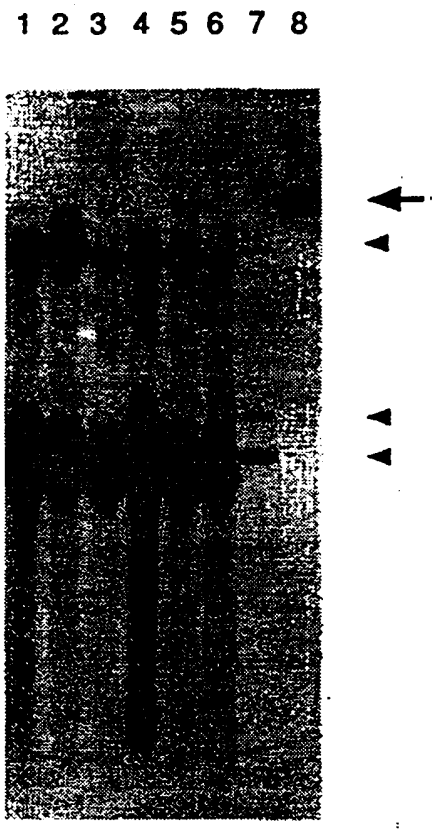

FIG. 3A

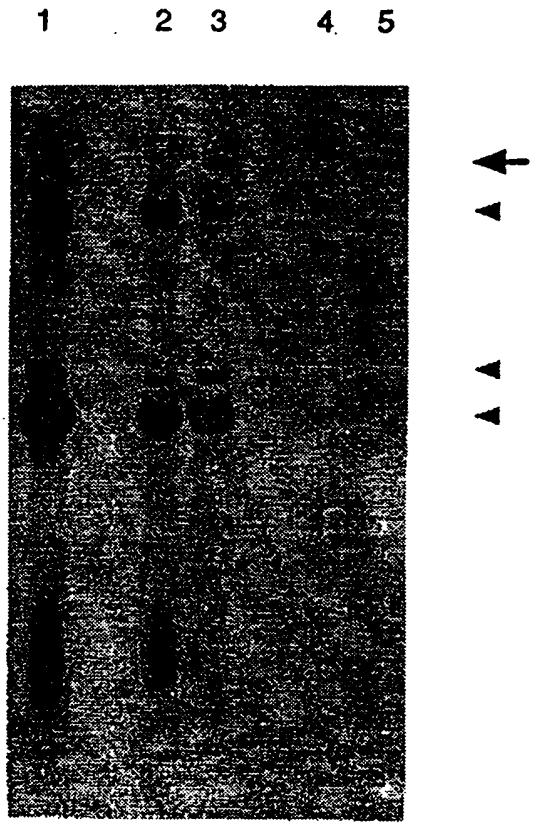

FIG. 3B

```
          -219   GTGAAGTGCTCAGAATGGGGCAGGATGTCACCTGGAATCAGCACTAAGTGATTCAGACTTTCCTTACTTTTAAATGTGCTGCTCTTCATTTCAAGATGC   -121

CGTTGCAGCTCTGATAAATGCAAACTGACAACCTTCAAGGCCACGACGGAGGGAAAATCATTGGTGCTTGGAGCATAGAAGACTGCCCTTCACAAAGGAAATCCCTGATTATTGTTTGAA     -1

ATGCTGAGGACGTTGCTGCGAAGGAGACTTTTTTCTTATCCCACCAAATACTACTTTATGGTTCTTGTTTTATCCCTAATCACCTTCTCCGTTTTAAGGATTCATCAAAAGCCTGAATTT    120
 M  L  R  T  L  L  R  R  R  L  F  S  Y  P  T  K  Y  Y  F  M  V  L  V  L  S  L  I  T  F  S  V  L  R  I  H  Q  K  P  E  F     40

GTAAGTGTCAGACACTTGGAGCTTGCTGGGGAGAATCCTAGTAGTGATATTAATTGCACCAAAGTTTTACAGGGTGATGTAAATGAAATCCAAAAGGTAAAGCTTGAGATCCTAACAGTG    240
 V  S  V  R  H  L  E  L  A  G  E  N  P  S  S  D  I  N  C  T  K  V  L  Q  G  D  V  N  E  I  Q  K  V  K  L  E  I  L  T  V     80
                                   *                 *

AAATTTAAAAAGCGCCCTCGGTGGACACCTGACGACTATATAAACATGACCAGTGACTGTTCTTCTTTCATCAAGAGACGCAAATATATTGTAGAACCCCTTAGTAAAGAAGAGGCGGAG    360
 K  F  K  K  R  P  R  W  T  P  D  D  Y  I  N  M  T  S  D  C  S  S  F  I  K  R  R  K  Y  I  V  E  P  L  S  K  E  E  A  E    120
                                         *

TTTCCAATAGCATATTCTATAGTGGTTCATCACAAGATTGAAATGCTTGACAGGCTGCTGAGGGCCATCTATATGCCTCAGAATTTCTATTGCGTTCATGTGGACACAAAATCCGAGGAT    480
 F  P  I  A  Y  S  I  V  V  H  H  K  I  E  M  L  D  R  L  L  R  A  I  Y  M  P  Q  N  F  Y  C  V  H  V  D  T  K  S  E  D    160

TCCTATTTAGCTGCAGTGATGGGCATCGCTTCCTGTTTTAGTAATGTCTTTGTGGCCAGCCGATTGGAGAGTGTGGTTTATGCATCGTGGAGCCGGGTTCAGGCTGACCTCAACTGCATG    600
 S  Y  L  A  A  V  M  G  I  A  S  C  F  S  N  V  F  V  A  S  R  L  E  S  V  V  Y  A  S  W  S  R  V  Q  A  D  L  N  C  M    200

AAGGATCTCTATGCAATGAGTGCAAACTGGAAGTACTTGATAAATCTTTGTGGTATGGATTTTCCCATTAAAACCAACCTAGAAATTGTCAGGAAGCTCAAGTTGTTAATGGGAGAAAAC    720
 K  D  L  Y  A  M  S  A  N  W  K  Y  L  I  N  L  C  G  M  D  F  P  I  K  T  N  L  E  I  V  R  K  L  K  L  L  M  G  E  N    240
```

FIG. 5A

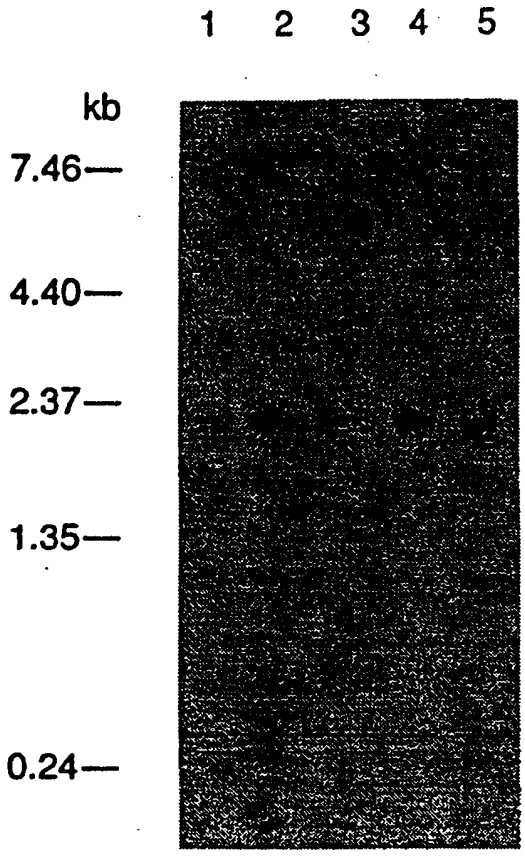

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/09303

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :Please See Extra Sheet.
US CL :435/6, 69.1, 193, 240.1, 252.3, 320.1; 530/395, 536/23.2, 23.5
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 69.1, 193, 240.1, 252.3, 320.1; 530/395, 536/23.2, 23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Medline, Embase, Pascal, Derwent World Patent Index
search terms: c2gnt, acetylglucosaminyltransferase, core 2, glucosaminyltransferase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Proceedings of the National Academy of Sciences USA, Vol. 86, issued February 1989, A. Pallant et al., "Characterization of cDNAs encoding human leukosialin and localization of the leukosialin gene to chromosomone 16", pages 1328-1332, see the entire document.	9-15
Y	Journal of Biological Chemistry, Vol. 266, No. 35, issued 15 December 1991, P. A. Ropp et al., "Mucin biosynthesis: Purification and characterization of a mucin beta-6N-acetylglucosaminyltransferase", pages 23863-23871, see the entire document.	1-8, 16-25

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be part of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 December 1993	29 DEC 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer ERIC GRIMES Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/09303

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

C07K 13/00; C12N 1/21, 5/16, 9/10, 15/12, 15/54, 15/63, 15/79; C12P 21/00; C12Q 1/68